Flow-resolution Enhancement in Electrophoretic NMR Using De-noising and Linear Prediction

Sunitha B. Thakur, Yinpeng Jin‡, Haihang Sun, Andrew F. Laine‡, and Qiuhong He
Department of Medical Physics and Radiology, Memorial Sloan-Kettering Cancer Center, NY;
‡Department Biomedical Engineering, Columbia University, NY

Abstract:

Detection of electrophoretic motion of ionic species using multi-dimensional Electrophoretic NMR (nD-ENMR) has demonstrated the potential to distinguish signals from two molecules in a solution mixture without their physical separation (1). Therefore, this technique may be applied for simultaneous structure determination of proteins and protein conformations, even during their biochemical interactions. Indeed, this has been achieved by introducing an additional dimension of electrophoretic mobility to the conventional multi-dimensional NMR by applying an external DC electric field. Consequently, the protein spectra are differently modulated by their electrophoretic mobilities in the electrophoretic flow dimension. Unfortunately, spectral resolution in the flow dimension has been limited by severe signal truncations due to the limited DC electric field available before onset of heating-induced convection. Linear prediction (2), which have been widely used for high-resolution spectral estimation from finite Fourier samples, have already been proposed to extend the truncated ENMR flow oscillation curves(3). However, we found that the spectral quality of linear prediction deteriorates as the spectral S/N decreases. To alleviate this problem, we have denoised the ENMR data using low pass filters prior to linear prediction. This technique has lead to improved resolution in the electrophoretic flow dimension. The approach was applied to analyze a 2D ENMR data matrix obtained from a mixture solution of two proteins ubiquitin and bovine serum albumin (BSA) in D_2O.

ENMR Separation of Co-existing Protein Signals in Solution:

When a DC electric field is applied to a solution of protein mixture, each protein component (ρ_i, i=1,2) moves with its characteristic electrophoretic mobility μ_i. The detectable NMR signal comes from both molecules:

$$M_{mix}(E_{dc}) = M_{p1}(E_{dc}) + M_{p2}(E_{dc}).$$

Where the signal of each protein component is modulated by a cosine factor:

$$M_{pi}(E_{dc}) = (M(0)/2)\exp[-D K^2(\tau_D - (\delta_1/3)) - 2\tau/T_2 - t_1/T_1]*\cos[(K E_{dc} \Delta)\mu_i].$$

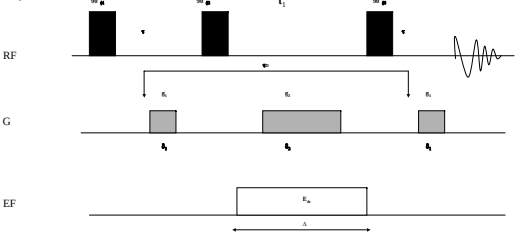

Fig 1: Pulse diagram of stimulated spin-echo ENMR

As an example, when an ENMR experiment (Fig.1) was carried out using stimulated echo pulse sequence where electric field pulses were applied to the two electrodes inserted into the sample solution in a U-shaped sample tube, a cosinusoidal amplitude modulation of the stimulated echo was introduced by the electrophoretic motion of molecules. The signals of the different protein components can be sorted according to the different oscillation frequencies of the cosinusoidal modulation functions.

Numerical Simulations

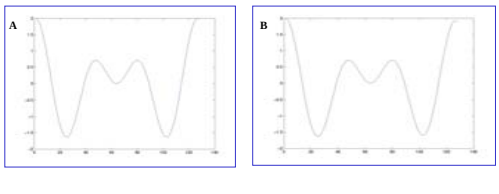

Fig. 2 (A) Sum of two synthetic cosinusoidal functions $\cos(w_1 t) + \cos(w_2 t)$ without noise (128 points). (B) The linear predicted curve of 128 points by considering the first 64 points for prediction, a curve nearly identical to the ideal data in (A).

The Truncation Limitation of ENMR:

In the ENMR experiments, we have frequently observed signal truncation in the flow dimension due to limited electric field available. This leads to poor flow separation of molecules having different electrophoretic mobilities.

Extrapolation of truncated data in Flow dimension

Linear Prediction (LP) is used to extrapolate the truncated time domain data. In these procedures it is assumed that signal is a sum of a series of data which are sampled at regular time intervals. In the linear prediction algorithms, each data point is expressed as a linear combination of M previous data points, where M is defined as the order of prediction coefficients. Linear prediction coefficients are evaluated to minimize the standard deviation of the predicted data points from the experimental results.

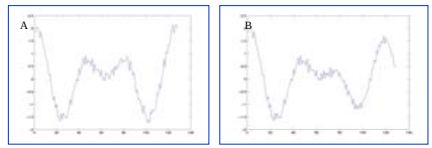

Fig. 3 (A) The previous synthetic curve from two cosinusoidal function $\cos(w_1 t) + \cos(w_2 t)$ was added with 40% of gaussian random noise. (B) The signal linear predicted from the first 64 points to 128 points has shown a severe signal distortion. The optimized linear prediction coefficients are 6. The prediction efficiency deteriorates as the S/N decreases.

Linear prediction works ideally in the absence of noise(Fig. 2). As S/N ratio decreases, however, the prediction efficiency reduces after certain number of data points (Fig. 3). In our ENMR data analysis, this problem can be alleviated by introducing a low-pass filter (4) to remove the noise (Fig. 5). The denoised data array was used as an input data to the linear prediction algorithm. The linear prediction coefficients are optimized using Marple's algorithm (5).

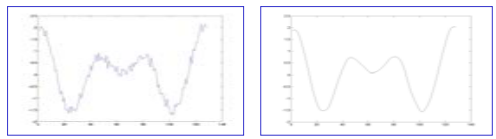

Fig. 4. A) Synthetic signal with 40% Gaussian noise (B) The signal extrapolated to 128 points using the first 64 points for prediction after signal denoising using the 4th order Butterworth filter. The signal distortion is minimized as compared with the data shown in Fig.3B.

Experimental Results:

We have applied the approach using the low-pass filtering procedure to remove the spectral noise in the flow dimension prior to linear prediction to analyze an experimental ENMR data matrix obtained from a protein mixture of 0.15mM BSA and 1mM ubiquitin with 26mM ethylenediamine in D_2O (pH=9.72) using a glass U-tube (1). The overlapping signals of the two protein components can be separated according to their electrophoretic mobilities (Fig. 6A). Experiments were carried out using Bruker-500 AMX spectrometer. A 2D ENMR data matrix (1024 x 21) was acquired with 21 steps incrementation of the electric field pulses in the flow dimension with the amplitude of the electric field E_{dc} changes from 0 to 40.08 V cm^{-1}. Fourier transformation in the chemical shift dimension displays electrophoretic interferogram. High spectral noise (Fig. 5A) was successfully removed using the low-pass filter(Fig.5B).

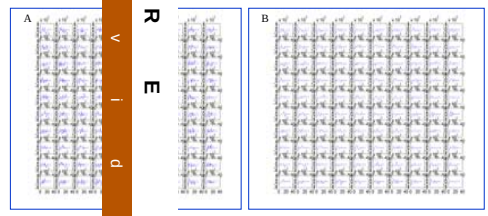

Fig. 5 (A) The
of BSA and U
shift dimension
4th order Butter

mns of 2-D ENMR data matrix from the protein mixture
flow dimension after Fourier transforming the chemical
esponding columns after removing the white noise using
filter.

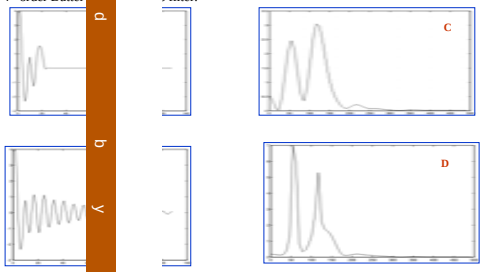

Fig. 6 (A) A D
21 steps of the
predicted to 64
coefficients op
due to the he
frequency dom

n taken from the 2D ENMR data matrix that is truncated after
incrementation. (B) The 21 time domain points were linear
first 16 points were taken for linear prediction with 6 LP
Marple's algorithm. (C) The signal decay in the time domain
nvection is translated into a line broadening factor in the
of the linear predicted signal improves signal resolution.

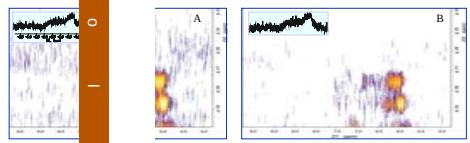

Fig. 7 Den
spectral fea
ENMR Spe
of the orig
proteins ar
ENMR data
using the sa
6B. There
for the spec

ar prediction in the flow dimension do not alter the NMR
emical shift dimension, as demonstrated by the 2D STE-
+Ubiquitin. (A) The 2D spectrum obtained after 2D FFT
ain data matrix. The overlapping signals of the two
the electrophoretic flow dimension. (B) The 2D STE-
ng and linear prediction treatment in the flow dimension,
cients optimized for a 1D flow column as shown in Fig.
stortions in the chemical shift dimension (See the inserts
of the 2D matrix).

Conclusio

The trunca
using a lin
remove th
applied to
signals of
the electro
multidime
determinat

horetic oscillation curves can be effectively extended
on algorithm after a low-pass filtering procedure to
oise in the flow dimension. The approach can be
gnal resolution in the flow dimension for resolving
otein components co-existing in solution according to
gration rates. The improvement of flow resolution in
MR is crucial for the simultaneous structure
ins.

Reference

1. He.Q e
ACS 120, 1341.

2. Press,
Computin
"Numerical Recipes in Fortran: the art of scientific
ge Univ. Press, 2nd Edition (1992).

3. Morris
hnson.C.S.Jr. (1992), JACS 114, 776; (1993), JMR
100A, 67.

R and W
ital Image Processing, Addison-Wesley, 2nd Edition
(1987). 4.

5. Marple
EEE Trans. on ASSP 28, 441.

6. He, Q
. JMR 141, 355.